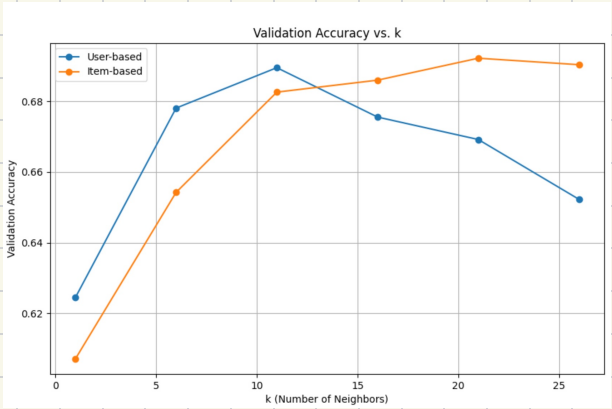
a, c)



a)

```
Validation Accuracy (User-based, k=1): 0.6244707874682472
Validation Accuracy (User-based, k=6): 0.6780976573525261
Validation Accuracy (User-based, k=11): 0.6895286480383855
Validation Accuracy (User-based, k=16): 0.6755574372001129
Validation Accuracy (User-based, k=21): 0.6692068868190799
Validation Accuracy (User-based, k=26): 0.6522720858029918
```

c)

```
Validation Accuracy (Item-based, k=1): 0.607112616426757
Validation Accuracy (Item-based, k=6): 0.6542478125882021
Validation Accuracy (Item-based, k=11): 0.6826136042901496
Validation Accuracy (Item-based, k=16): 0.6860005644933672
Validation Accuracy (Item-based, k=21): 0.6922099915325995
Validation Accuracy (Item-based, k=26): 0.69037538808919
```

b, c)

```
Best k (User-based): 11
Best k (Item-based): 21
```

b, c)

```
Test Accuracy on k*:
Validation Accuracy (User-based, k=11): 0.6841659610499576
Validation Accuracy (Item-based, k=21): 0.6816257408975445
```

is on test_data

c) underlying assumption is if question A has the same correct/incorrect responses from other students as question B, then the correctness of a specific student's response to A will match their response to B.

d) according to b/c, the test performance of user-based performs a bit better.

e) 1. scalability, since KNN needs to compute pairwise distance is expensive the more data there is

2. false assumptions: KNN assumes similarity between questions/users captures the relationship but it might be false and more complex, would be inaccurate if these assumptions are not true